

RESEARCH ARTICLE

Open Access



Joint Bayesian modeling of time to malaria and mosquito abundance in Ethiopia

Denekew Bitew Belay^{1*} , Yehenew Getachew Kifle², Ayele Taye Goshu¹, Jon Michael Gran³, Delenasaw Yewhalaw⁴, Luc Duchateau⁵ and Arnoldo Frigessi³

Abstract

Background: This paper studies the effect of mosquito abundance and malaria incidence in the last 3 weeks, and their interaction, on the hazard of time to malaria in a previously studied cohort of children in Ethiopia.

Methods: We model the mosquito abundance and time to malaria data jointly in a Bayesian framework.

Results: We found that the interaction of mosquito abundance and incidence plays a prominent role on malaria risk. We quantify and compare relative risks of various factors, and determine the predominant role of the interaction between incidence and mosquito abundance in describing malaria risk. Seasonal rain patterns, distance to a water source of the households, temperature and relative humidity are all significant in explaining mosquito abundance, and through this affect malaria risk.

Conclusion: Analyzing jointly the time to malaria data and the mosquito abundance allows a precise comparison of factors affecting the spread of malaria. The effect of the interaction between mosquito abundances and local presence of malaria parasites has an important effect on the hazard of time to malaria, beyond abundance alone. Each additional one km away from the dam gives an average reduction of malaria relative risk of 5.7%. The importance of the interaction between abundance and incidence leads to the hypothesis that preventive intervention could advantageously target the infectious population, in addition to mosquito control, which is the typical intervention today.

Keywords: Mosquito abundance, Time to malaria, MCMC, Abundance and incidence interaction, Bayesian inference

Background

Malaria has an estimated incidence of more than 300 million new cases every year world wide. About one million people die each year due to malaria, of which most of them are in sub-Saharan Africa [1–4]. Approximately 90% of the malaria cases are related to environmental factors [5]. In Ethiopia three-fourths of the land below 2000 meters of altitude is malarious, with two-thirds of the country's population at risk of malaria infection. In Ethiopia alone there is an average of 5 million cases a year, causing 70,000 deaths each year and accounting for 17% of outpatient visits to health institutions [6]. It has been argued that economic growth is dependent on two key factors,

namely the timely arrival of seasonal rain fall and malaria epidemics [7–9].

The establishment and operation of water resource development projects represents an important risk factor, since dams and irrigation schemes transform ecosystems and can substantially change the nature of malaria risk proximal to their location. There is a substantial body of literature documenting the increases in malaria incidence as a consequence of such projects [3]. Ethiopia has recently constructed a large number of dams to produce electricity and for irrigation [10, 11]. Even though dams give economic benefits, they can increase the survival, density and distribution of disease vectors transmitting parasites such as malaria, by providing appropriate habitats. In the region of Ethiopia where this study has been conducted, *Anopheles gambiae sensu lato* is the vector of the malaria parasites, *Plasmodium falciparum* and *Plasmodium vivax*, which occur in the blood of humans

*Correspondence: denekew.t.h@gmail.com

¹School of Mathematical and Statistical Sciences, College of Natural and Computational Science, Hawassa University, Hawassa, Ethiopia
Full list of author information is available at the end of the article

who are affected by malaria, during clinical episodes of the disease.

The data in this study originate from a project which aimed to assess the effect on malaria incidence [10] of the construction of a mega hydropower dam in Gilgel Gibe in southwest Ethiopia. A cohort of healthy children was followed over two years, recording the first malaria episode for each cohort of children. Important covariates are the shortest distance of households to the dam shore, seasonal precipitations, temperature and relative humidity. Mosquito counts were also collected in each village, see below for details.

In this paper we explore *jointly* the association between longitudinal measurement of mosquito abundance and time to malaria in the same cohort of children. We investigate the effect of distance, seasonal precipitations, temperature, structure of the house and relative humidity in a mixed effect model for the measured mosquito abundance. This allows us to define a latent variable of mosquito abundance in every village which is more regular than the measured abundance. These latent variables enter into a proportional hazard model for the time to malaria data, together with their interactions with an approximated measure of the time varying malaria incidence in each village. The two models are estimated jointly using an iterative procedure (Markov Chain Monte Carlo) developed and implemented in [12]. Our approach follows a Bayesian version of the joint modeling of longitudinal and time to event data [13].

We investigate whether the distances of households to the dam and the seasonal factors are significantly associated to mosquito abundance. We study the effect of the regularised mosquito abundance in the malaria hazard model, and quantify the role of its interaction with the measure of malaria incidence village-wise on malaria risk.

Previously, Getachew et al. [10] used two separate models, a mixed Poisson regression model for the mosquito abundance and a frailty model for the time to malaria. They found that distance of households to the dam was not significantly associated to malaria incidence, while seasonal effects were significantly associated with malaria incidence. In the paper [11] the authors investigated further the effect of distance of households from the dam on *Anopheles* mosquito abundance. They found that the abundance was varying significantly between seasons and was higher in villages near the dam.

Gilgel Gibe data

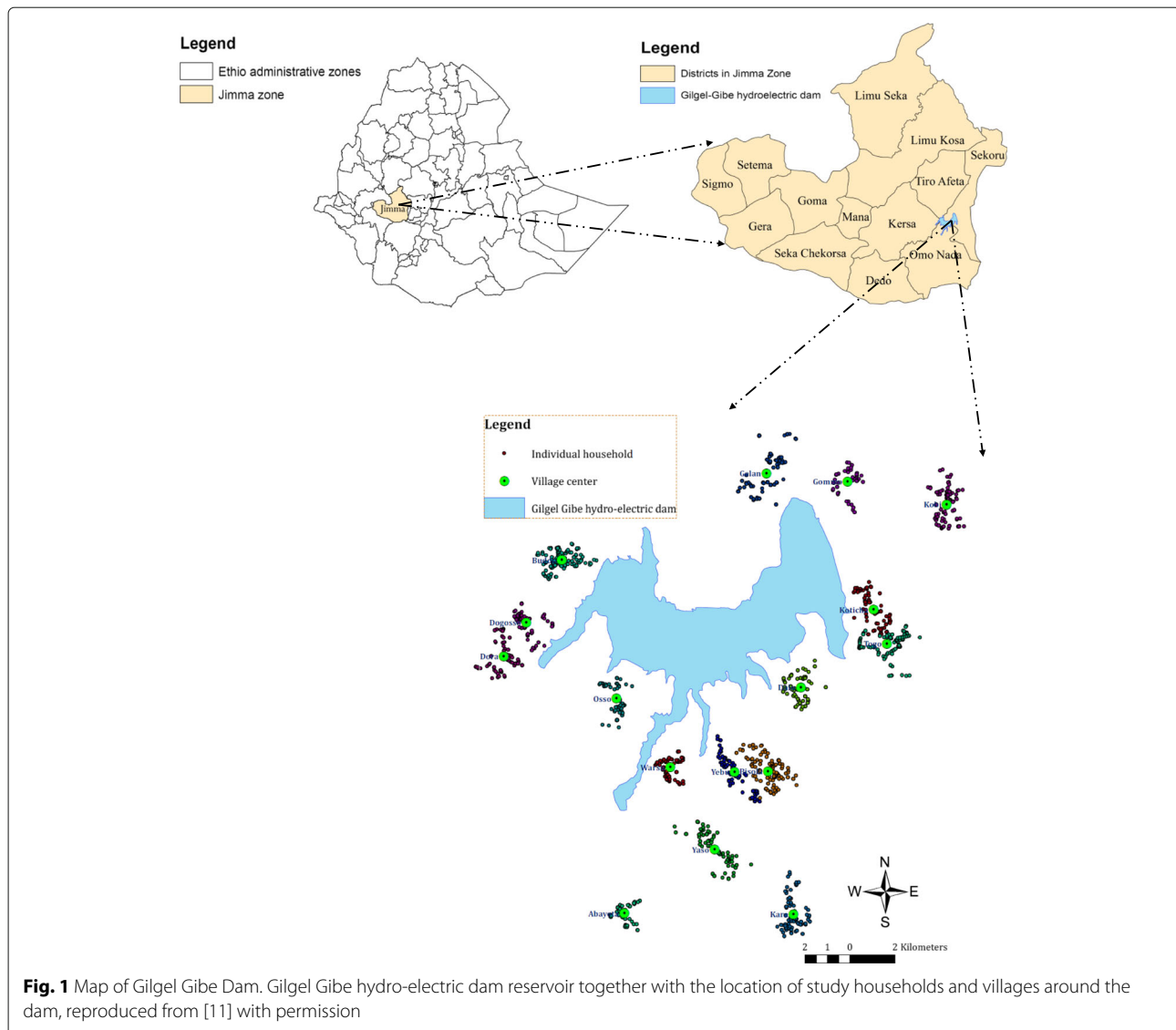
In this paper we analyse the data previously studied in [10]. The study area is in southwest Ethiopia, around the Gilgel Gibe hydro-electric dam, which created an artificial lake. Sixteen villages were part of the study, in the four districts Sekoru, Tiro afeta, Omo nada and Kersa. The Gilgel Gibe hydro-electric dam area, together with the

location of the study households in sixteen villages around the dam are shown in Fig. 1, taken from [11]. The sixteen villages were selected at various distances from the closest dam shore, ranging from 0.265 to 9.056 km. The study period was 23 months long, between 2008 and 2010. In each village on average 130 households were sampled, and in each household one healthy child of age below ten years was randomly chosen and followed from study start. Twenty-nine (1.4%) children died due to various reasons (possibly also malaria related) and 15 (0.72%) migrated elsewhere during the study. Of the migrant children, two children were tracked back and 13 were lost to follow up and excluded from the analysis.

Each child in the cohort of 2040 children was visited weekly during the 23 months of the study and each child was monitored by a trained data collector, who measured his or her body temperature. Parents were asked if the child had signs of fever in the preceding week. In either case, blood from the finger of the child was collected and sent for laboratory analysis to Jimma University referral hospital. There, a blood test was performed to establish if the child had contracted malaria or not. The presence of laboratory confirmed malaria fever is defined as the event in this study. Children who contracted malaria were treated as necessary. Censoring is only related to children who were event free at the end of the study. Age and gender of each child were recorded, together with other variables related to the household. The distance of each household in the study to the closest dam shore was calculated using GPS coordinates.

In the study, an estimate of the prevalence of malaria transmitting mosquitoes has been obtained in the following way. In each village, two houses, one in the center and the other in the periphery of each village, were selected. In each such house mosquitoes were trapped and counted during one night every month using light trap catches from 6pm to 6am. The counts range from 0 mosquitoes to 280 per night in our data. In order to assign a single mosquito abundance estimate per village, we took the average of the two counts per village. This estimated abundance, considered as the exposure to mosquitoes, is therefore the same for all the houses in a village within each month.

Seasonal variability in mosquito abundance is well known, due to varying precipitation, humidity and temperature. Daily rainfall (in mm), monthly average temperature (in °C) and monthly average relative humidity (in percentage) for the whole area have been obtained from the south-western branch regional office of the Ethiopian Meteorological Agency, and are shared by all villages. Finally the way houses are build can also play a role, in particular if they are traditionally build with mud and stick or thatch walls or with corrugated iron. See [4] and [10] for more details.



Methods

We first perform a simple exploratory data analysis, to investigate visually relations between new malaria cases, rain and mosquito abundance data. We then study the association between the longitudinal measurements and the event of interest through a Bayesian joint model. The mosquito abundance represents the longitudinal component, while the time to malaria is the survival part of the joint model. Further explanations on the derivation of the joint models can be found in the book [13].

Longitudinal component for anopheles mosquitoes abundance

The time unit in our models is the week t . Let $x_i(t)$ be the measured abundance of mosquitoes at week t for child i , as measured in her/his village. As abundance data were measured once every month, $x_i(t)$ is constant over four

week periods. Let $k(i) \in 1, 2, \dots, 16$ indicate the village of child i . We assume

$$x_i(t) = N_{k(i)}(t) + \epsilon_i(t), \quad (1)$$

where $N_{k(i)}(t)$ is a real valued latent process describing a regularized version of the mosquito abundance at time t in the village $k(i)$ of child i , once a measurement error $\epsilon_i(t) \sim \mathcal{N}(0, \sigma^2)$, independent of the other variables, is subtracted from the actual counts.

Next we define a useful weather related covariate. In Ethiopia there are three yearly seasons, namely Bega (dry season), Kiremt (long rainy season) and Belg (short rainy season), based on the magnitude and distribution of rain-fall across the 12 months of the year [11]. In [10] these three seasons were identified for the two years and for

the region of southwest Ethiopia in our study as the following seven climate periods:

$$\text{period}(t) = \begin{cases} 1, & \text{when } t \in [\text{weeks in July}] \text{ in 2008} \\ 2, & \text{when } t \in [\text{weeks from August to November}] \text{ in 2008} \\ 3, & \text{when } t \in [\text{weeks from December to March}] \text{ in 2009} \\ 4, & \text{when } t \in [\text{weeks from April to July}] \text{ in 2009} \\ 5, & \text{when } t \in [\text{weeks from August to November}] \text{ in 2009} \\ 6, & \text{when } t \in [\text{weeks from December to March}] \text{ in 2010} \\ 7, & \text{when } t \in [\text{weeks from April to June}] \text{ in 2010.} \end{cases} \quad (2)$$

The first and last periods are shorter than others, which last four months. See also [11] about this definition of the seasonal periods.

The daily rainfall (in mm/day) $\text{precip}(s)$ is available for each day s , but we found it useful to use smoothed versions. In our model the best fit, based on comparison of the Deviance Information Criterion (DIC) [14], was obtained when averaging over the above climate periods: we define

$$\text{rain}(t) = \frac{1}{\text{number of days in period}(t)} \sum_{s \in \text{period}(t)} \text{precip}(s)$$

In order to smooth the step function $\text{rain}(t)$, we approximated it with a natural cubic spline with three degrees of freedom and denote as $S_1(\text{rain}(t))$ and $S_2(\text{rain}(t))$ the two B-spline base function components of $\text{rain}(t)$, see [12] and [15] for details. In our data there is one global measure of rainfall shared by all children in all villages, which explains the seasonal pattern of the mosquito abundance. Let $\text{temp}(t)$ be the monthly average temperature and $\text{humid}(t)$ the monthly average humidity of the month of week t . We denote by corrugate_i the variable taking value 1 if the house of child i had a corrugate iron roof, 0 otherwise. Finally, let $\text{dist}_{k(i)}$ be the distance from the center of the village $k(i)$ of child i to the closest dam shore.

For the latent abundance $N_{k(i)}(t)$ we assume the following random effect model:

$$N_{k(i)}(t) = \beta_0 + \beta_1 S_1(\text{rain}(t)) + \beta_2 S_2(\text{rain}(t)) + \beta_3 \text{dist}_{k(i)} + \beta_4 \text{temp}(t) + \beta_5 \text{humid}(t) + \beta_6 \text{corrugate}_i + w_{i0} + w_{i1} S_1(\text{rain}(t)) + w_{i2} S_2(\text{rain}(t)), \quad (3)$$

where $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)$ is the vector of fixed effects and $w_i = (w_{i0}, w_{i1}, w_{i2})$ is the vector of random coefficients with three component. We assume that w_i is a priori normally distributed with zero means and covariance matrix D

$$w_i = (w_{i0}, w_{i1}, w_{i2}) \sim \mathcal{N}_3(0, D),$$

where D is specified below. We also tested a model without the distance as covariate.

Time to event model component for malaria contraction

Let T_i denote the time (in weeks) to malaria infection for child i , which can be censored at the study end point in week 100. We used a Cox proportional hazard model [16], where the hazard of child i at time t is

$$h_i(t) = h_0(t) \exp \{ \theta_1 \text{age}_i + \theta_2 \text{gender}_i + \alpha_1 N_{k(i)}(t) + \alpha_2 N_{k(i)}(t) I_{k(i)}(t) \}. \quad (4)$$

Here age_i is the age of child i at week one, gender_i the gender of the child and the covariate $I_{k(i)}(t)$ is the total number of new malaria cases in the village $k(i)$ of child i in the last three weeks before and including week t . This is a proxy for a local incidence of malaria because the infectious period of malaria infected children lasts approximately 3 weeks. The term $N_{k(i)}(t) I_{k(i)}(t)$ represents the interaction between mosquito abundance and incidence. We also tested separately the incidence with a two weeks windows. As suggested in [12], we model the baseline logarithm of the hazard $h_0(t)$ with a B-spline with a quite large number of knots, 15 in our analysis, in order to allow for flexibility. Knots are equally spaced according to the percentiles of the observed event times. We assume

$$\log h_0(t) = \phi_0 + \sum_{z=1}^{15} \phi_z B_z(t, \nu)$$

where $B_z(t, \nu)$ denotes the z^{th} basis function of a B-spline with knots $\nu = \nu_1, \nu_2, \dots, \nu_{15}$ and ϕ is the vector of all spline coefficients, which are then penalised for smoothness, penalising with the integral of the second derivative of the fitted log hazard, as explained in [12], using a penalty coefficient denoted by λ .

Priors

The prior probability distribution for the fixed effects $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)$ of the linear mixed effects model is assumed to be normal with means equal to zero and a seven by seven precision matrix, i.e. the inverse of the covariance matrix, with diagonal elements equal to 0.01 and zero otherwise.

The prior probability distribution for the regression coefficients (θ_1, θ_2) of the survival model in (4) is assumed normal with mean zero and a 2 by 2 precision matrix with diagonal 0.01 and zero otherwise.

A normal probability distribution prior is also assumed for the association parameters (α_1, α_2) in the survival model, again with means equal to zero and a 2 by 2 precision matrix with diagonal elements 0.1 and zero otherwise.

For the parameters of the log baseline hazard function we follow [17] and assume for the coefficients ϕ of the

B-spline an improper prior and for the smoothing parameter λ a Gamma hyperprior as follows:

$$P(\phi|\lambda) \propto \lambda^{\frac{r(A)}{2}} \exp\left(-\frac{\lambda}{2}\phi^T A \phi\right)$$

where $A = \Delta_2^T \Delta_2$, where Δ_2 is the order two difference penalty matrix and $r(A)$ denotes the rank of A . For the smoothing parameter λ we assume a $\text{Gamma}(1, 0.005)$ prior, which leads to a proper posterior for ϕ ([17]). More detailed information can be found in [12, 15].

The prior for the precision matrix D^{-1} of the the random effects in the linear mixed model is a Wishart distribution, because it is the conjugate prior for the precision matrix in Gaussian models [18]. For the hyperprior we performed a preliminary run of the *lme*-function in R for model (1) and (3) to obtain an estimate of D^{-1} , which we used to fix the scale matrix in the prior as follows:

$$D^{-1} \sim \text{Wish}\left(\begin{bmatrix} 53903.4 & 1980.7 & -1096.7 \\ 1980.7 & 75.9 & -42.1 \\ -1096.7 & -42.1 & 25.6 \end{bmatrix}, 3\right).$$

We also used this scale matrix as starting point of the MCMC, to speed up convergence.

A similar technique was used for the prior probability distribution of the precision parameter $(\sigma^2)^{-1}$. This was assumed to be Gamma distributed with hyper-parameters obtained by running the *lme*-function for the mixed effect model (1) and (3). The idea of using a preliminary run in order to fix hyperparameters in the prior in an empirical Bayesian fashion, has been suggested in [12], where also a further general discussion on the choice of priors in the Bayesian joint model is given. See also [19].

Implementation

We used the R-package JMBayes [12] for our analysis. It implements a full scale MCMC algorithm to sample all parameters in the posterior model. Briefly, the first step is to run the linear mixed effect model on its own and the Cox proportional hazard model with only gender and age as covariates, in order to obtain good starting values for most parameters in the MCMC. Then a Metropolis Hasting algorithm is started, with random walk proposals. Convergence of the MCMC was tested using trace plots and checking robustness of estimates when repeating the runs with different random starting points. We concluded that it was appropriate to run 200,000 iterations in our analysis, after a burn-in of 120,000 iterations. For further details on the algorithm see [12]. Credible intervals are used to describe the variability of the posterior around the posterior mean.

Results

In Fig. 2, we plot the total new malaria cases in the last three weeks in all villages together (left axis). This shows a clear association between our measure of incidence and the seasonal rain pattern (right axis). In the main rainy season (coded red in the x-axis), malaria incidence is higher. In the short rains season (black) the incidence is higher than in dry season (green) when incidence is low.

In Fig. 3 we show the total mosquito abundance (left axis) in all villages as a function of time and average daily rain (right axis). We see that abundance is low in the dry season and builds up during the short rainy season (black) and main rainy season (red). The highest abundance comes typically early in the rainy season, and then it decreases, which might be due to heavy rain flooding mosquito breeding habitats.

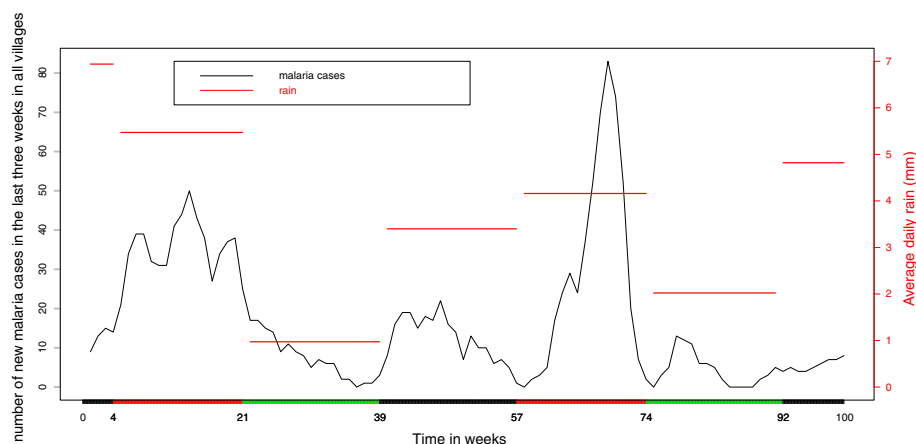


Fig. 2 Malaria cases in the last three weeks Vs rain. Total number of malaria cases in the past three weeks before week t in all villages (left axis) and average of daily rain(mm) (right axis). On the x-axis the colours identify the three seasons: dry season (green), long rainy season (red), short rainy season (black)

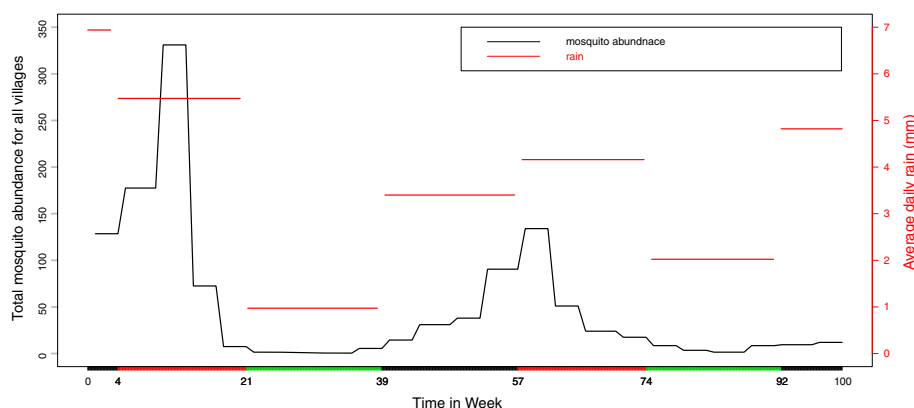


Fig. 3 Total mosquito abundance in the past four weeks Vs rain. Total mosquito abundance in the past four weeks before week t in all villages (left axis) and average of daily rain (mm) (right axis). On the x-axis the colours identify the three seasons: dry season (green), long rainy season (red), short rainy season (black)

Based on the Bayesian joint model we obtain the posterior mean of the parameters in the model with 95% credible intervals, as given in Table 1. The association parameters α_1 and α_2 are different from zero with high posterior probability, as the 95% credible interval does not contain zero. This confirms the utility of the joint

Table 1 Parameter estimates and 95% credible intervals for the joint model

	Parameter	Posterior mean	2.5%	97.5%
Abundance model				
Intercept	β_0	3.12	2.89	3.36
$S_1(rain(t))$	β_1	5.99	5.72	6.26
$S_2(rain(t))$	β_2	0.67	0.59	0.74
Distance	β_3	-0.13	-0.17	-0.11
Temperature	β_4	-0.15	-0.16	-0.15
Relative humidity	β_5	0.0004	0.0001	0.0006
Corrugate roof	β_6	0.05	-0.07	0.17
Measurement error	σ	0.69	0.69	0.70
Time to event model				
Age	θ_1	0.01	-0.03	0.05
Gender	θ_2	-0.05	-0.21	0.12
Association main effect	α_1	0.14	0.06	0.21
Association interaction	α_2	0.31	0.20	0.41
Hyper-parameters				
Penalty	λ	0.0031	0.0014	0.0055
Random effect covariance	$D_{1,1}$	26.96	25.37	28.70
Random effect covariance	$D_{2,1}$	0.75	0.35	1.16
Random effect covariance	$D_{3,1}$	-0.80	-1.07	-0.53
Random effect covariance	$D_{2,2}$	3.05	2.84	3.27
Random effect covariance	$D_{3,2}$	0.82	0.71	0.93
Random effect covariance	$D_{3,3}$	1.40	1.31	1.50
DIC		398866.1		

D_{ij} denotes the ij -element of the covariance matrix for the random effects. We use a three week window to define the incidence $I_{k(i)}(t)$

model for the present data. The coefficient α_1 is the effect of mosquito abundance on the malaria hazard and has a posterior mean of 0.14, with 95% credible interval (0.06, 0.21). For each unit increase in abundance, assuming all other variables constant, the hazard increases by a factor of $\exp(0.14) = 1.15$. The coefficient α_2 controls the effect of the interaction between abundance and local incidence (defined by $I_{k(i)}(t)$) and has a posterior mean equal to 0.31 (0.20, 0.41). This indicates that the interaction $N_{k(i)}(t)I_{k(i)}(t)$ has an important effect on the hazard of malaria, beyond abundance $N_{k(i)}(t)$ alone. This is of course reasonable, as both the vector and the parasite must be present together for malaria to occur and spread. As age and gender are not significant (as found in previous studies), the posterior expected hazard can be assumed to be proportional to

$$\text{posterior expected hazard} \propto \exp \{0.14N_{k(i)}(t) + 0.31N_{k(i)}(t)I_{k(i)}(t)\}. \quad (5)$$

This shows that the interaction between the mosquito abundance $N_{k(i)}(t)$ and the local presence of the malaria parasite, for which $I_{k(i)}(t)$ is a proxy, is the most important factor. The quantification of this effect is interesting: the hazard increases by a factor of $\exp(0.14) = 1.15$ per unit increase of the abundance because of the first term in (5), while it increases by $\exp(0.31 \times I_{k(i)}(t))$ because of the second term in (5). As $I_{k(i)}(t)$ takes values between 0 and 12 across all villages, with mean 1.01, this relative risk is on average $\exp(0.31 \times 1.01) = 1.37$, but can reach values as high as $\exp(0.31 \times 12) = 41.26$.

The distance to the dam has a negative effect on abundance: the mosquito abundance decreases on average by 0.13 per kilometer away from water keeping other covariates constant. Consequently, the average hazard of time to malaria can be computed as $\exp(0.14 \times (-0.13) + 0.31 \times$

$(-0.13) \times 1.01 = 0.943$, where we used 1.01 which is the average local incidence in the last three weeks across all villages. This indicates an average reduction of 5.7% of the malaria relative risk per km away from the dam.

In Fig. 4, we plot the log posterior expected risk $\alpha_1 + \alpha_2 I_{k(i)}(t) = 0.14 + 0.31 I_{k(i)}(t)$ per unit mosquito abundance in all villages as a function of time. We can see that there is a seasonal variation in the risk of malaria infection per unit mosquito abundance in all villages. In particular it is higher in the second and the fifth periods, corresponding to the two main rainy seasons in the two years of the study. Beside this, the short rain seasons (Belg) from April to July also show a relatively higher risk of malaria. We also notice in Table 1 that the intercept is significant. This means that our covariates do not explain the mosquito abundance in full and that there is an overall abundance, present constantly and independently of meteorological covariates and distance. To a large extent, this is probably attributable to the fact that our covariates come with large measurement error.

Figure 5 shows the log posterior expected risk of malaria for six typical villages. We see that in all villages there is a high risk of malaria in the two rainy seasons, but there is a higher variability in the effect of the season between villages. We found that there is more dependence on season than on distance as also discussed in [11]. When the temperature increases, mosquito abundance has a tendency to decrease, similarly with the findings in [11]. Humidity is positive correlated to mosquito abundance, although the effect is minor.

Next we test if a reduction of the weather related covariates would lead to a better model. In Table 2, we use only the rain covariate and not temperature and humidity, nor roof construction. We use the Deviance Information Criterion (DIC) [14] to compare models. The DIC is larger,

favouring the use of all covariates. We also tried to drop the distance to the dam as covariate, see Table 3. Again the DIC indicates that the model including the distance and the weather covariates is to be preferred.

To study robustness of our results with respect to the definition of the incidence, we report in Table 4, the results for the full model with all three weather related covariates, roof construction and the distance, when we only change the definition of $I_{k(i)}(t)$ to include only the last two weeks (instead of three). Estimates are in practice unchanged, with a slightly larger effect of the distance. The DIC is essentially unchanged, with a slight preference for the three weeks definition of the village-wise incidence. Finally in Table 5, we tested the model with the same definition of incidence, but using only rain as weather covariate and distance. The DIC is again larger.

The role of distance has been discussed before in [10]. Table 3 reports the results when we use only seasonal rain as covariate in the abundance model (3). Here we use three weeks accumulated cases for $I_{k(i)}(t)$. The estimates of the remaining parameters are in practice unchanged, which shows that the distance (significant in Table 1) helps explain the abundance component in addition to the other covariates. The main difference between Tables 3 and 1 is that the intercept in the abundance model (3) does not stay away from zero with high posterior probability when the distance covariate is dropped, and the random effect covariances are also smaller. This shows, in regularisation of mosquito abundances using model (3), there is more subject variability in Table 1 than in Table 3, when we use only seasonal rain.

Discussion

In this paper we propose the use of a Bayesian joint model to integrate time to malaria and mosquito abundance

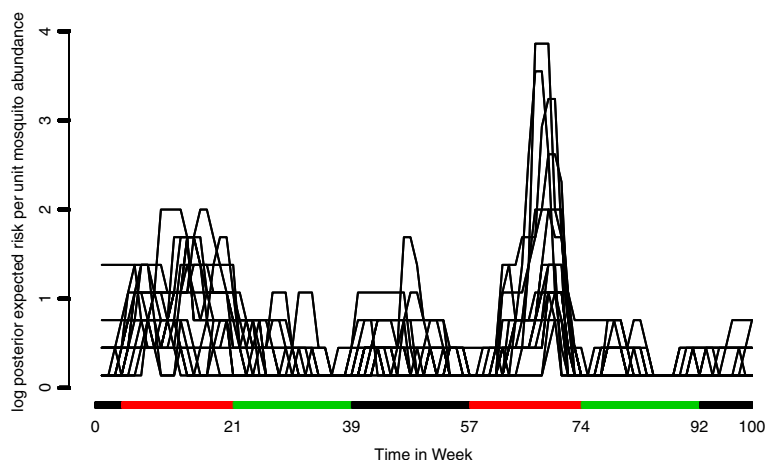


Fig. 4 Log posterior expected risk for all villages. The log posterior expected risk of malaria infection for the children in each village per unit increase of mosquito abundance. On the x-axis the colours identify the three seasons: dry season (green), long rainy season (red), short rainy season (black)

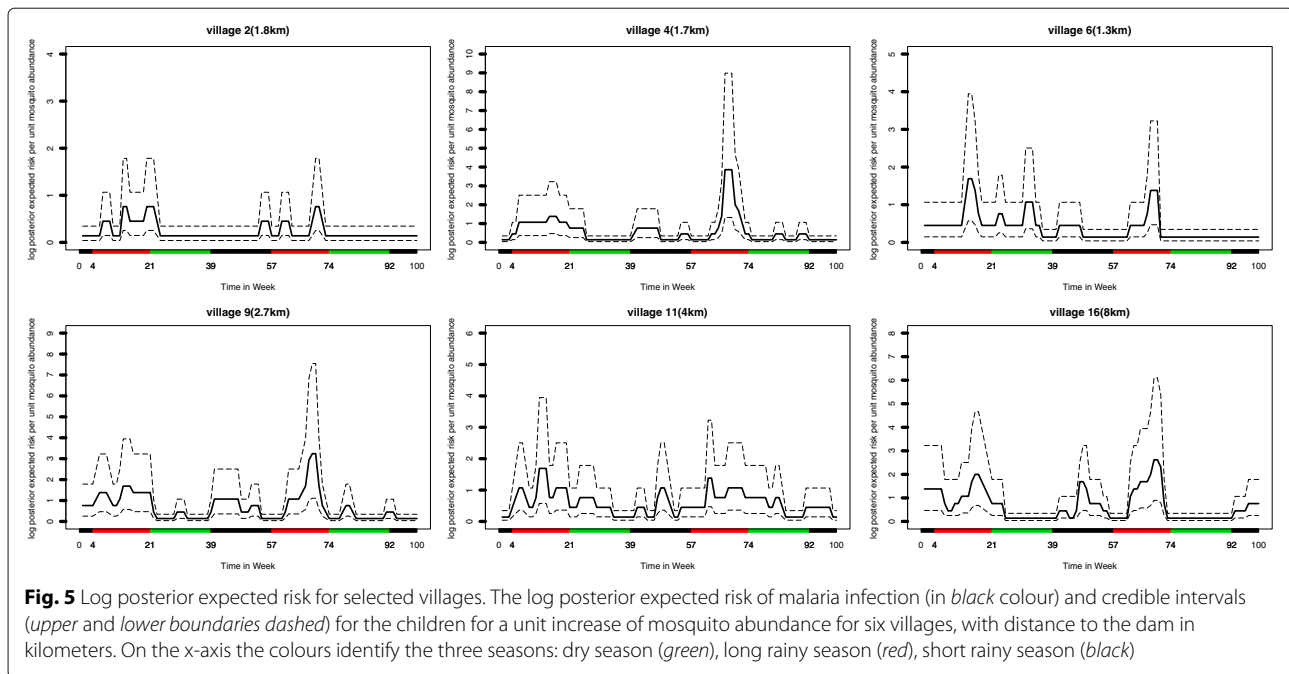


Table 2 Parameter estimates and 95% credible intervals for the joint model

	Parameter	Posterior mean	2.5%	97.5%
Abundance model				
Intercept	β_0	0.66	0.40	0.91
$S_1(rain(t))$	β_1	4.40	4.18	4.61
$S_2(rain(t))$	β_2	0.92	0.85	0.99
Distance	β_3	-0.19	-0.23	-0.12
Measurement error	σ	0.71	0.70	0.71
Time to event model				
Age	θ_1	0.01	-0.03	0.05
Gender	θ_2	-0.04	-0.21	0.12
Association main effect	α_1	0.12	0.04	0.19
Association interaction	α_2	0.27	0.16	0.37
Hyper-parameters				
Penalty	λ	0.005	0.002	0.008
Random effect covariance	$D_{1,1}$	24.22	22.78	25.77
Random effect covariance	$D_{2,1}$	0.94	0.60	1.27
Random effect covariance	$D_{3,1}$	-0.73	-0.99	-0.47
Random effect covariance	$D_{2,2}$	2.16	2.01	2.31
Random effect covariance	$D_{3,2}$	0.81	0.72	0.91
Random effect covariance	$D_{3,3}$	1.41	1.32	1.51
DIC		403478.6		

D_{ij} denote the ij -element of the covariance matrix for the random effects. Here rain is the only weather related covariate

Table 3 Parameter estimates and 95% credible intervals for the joint model

	Parameter	Posterior mean	2.5%	97.5%
Abundance model				
Intercept	β_0	0.18	-0.03	0.39
$S_1(rain(t))$	β_1	4.37	4.16	4.59
$S_2(rain(t))$	β_2	0.92	0.85	0.98
Measurement error	σ	0.71	0.70	0.71
Time to event model				
Age	θ_1	0.01	-0.03	0.05
Gender	θ_2	-0.05	-0.22	0.12
Association main effect	α_1	0.12	0.04	0.19
Association interaction	α_2	0.26	0.16	0.36
Hyper-parameters				
Penalty	λ	0.003	0.002	0.006
Random effect covariance	$D_{1,1}$	23.55	22.15	25.05
Random effect covariance	$D_{2,1}$	0.58	0.26	0.89
Random effect covariance	$D_{3,1}$	-0.73	-0.99	-0.49
Random effect covariance	$D_{2,2}$	2.16	2.01	2.31
Random effect covariance	$D_{3,2}$	0.81	0.72	0.91
Random effect covariance	$D_{3,3}$	1.42	1.32	1.52
DIC		403452.2		

D_{ij} denotes the ij -element of the covariance matrix for the random effects. We use a three week window to define the incidence $I_{k(t)}(t)$. In this run, the distance is not included in the model

Table 4 Parameter estimates and 95% credible intervals for the joint model

	Parameter	Posterior mean	2.5%	97.5%
Abundance model				
Intercept	β_0	3.28	3.04	3.52
$S_1(rain(t))$	β_1	6	5.73	6.27
$S_2(rain(t))$	β_2	0.67	0.59	0.74
Distance	β_3	-0.19	-0.20	-0.17
Temperature	β_4	-0.15	-0.16	-0.15
Relative humidity	β_5	0.0004	0.0002	0.0006
Corrugate roof	β_6	0.05	-0.07	0.17
Measurement error	σ	0.69	0.69	0.70
Time to event model				
Age	θ_1	0.01	-0.03	0.05
Gender	θ_2	-0.04	-0.21	0.12
Association main effect	α_1	0.14	0.07	0.21
Association interaction	α_2	0.31	0.20	0.41
Hyper-parameters				
Penalty	λ	0.0038	0.0017	0.0069
Random effect covariance	$D_{1,1}$	27.10	25.46	28.82
Random effect covariance	$D_{2,1}$	0.92	0.51	1.33
Random effect covariance	$D_{3,1}$	-0.82	-1.10	-0.55
Random effect covariance	$D_{2,2}$	3.05	2.85	3.27
Random effect covariance	$D_{3,2}$	0.81	0.71	0.92
Random effect covariance	$D_{3,3}$	1.40	1.31	1.50
DIC		398876		

D_{ij} denotes the ij -element of the covariance matrix for the random effects. We use a two week window to define the incidence $I_{k(t)}(t)$

longitudinal data. Our model allows to represent the causal flow in a natural way: distance to the water, seasonal precipitations and other covariates induce mosquito abundance, which in turn by interacting with the presence of malaria parasites, leads to new malaria cases. We found that the two parameters α_1 and α_2 are estimated significantly away from zero. This indicates that the two component of the joint model (the longitudinal and the time to event) are linked, justifying the importance of using a joint model approach.

The estimated hazard (as in Fig. 4), follows the seasonal pattern. The local precipitation, when regularised into appropriate seasons and expressed as average daily rain, is associated with new malaria cases recorded in the last three weeks and accumulated across villages. Also, our measure of mosquito abundance is correlated to average seasonal daily precipitation, with abundance peaking early in the main rainy season. This confirms the finding in [11] that observed that *Plasmodium falciparum* malaria

Table 5 Parameter estimates and 95% credible intervals for the joint model

	Parameter	Posterior mean	2.5%	97.5%
Abundance model				
Intercept	β_0	0.67	0.42	0.92
$S_1(rain(t))$	β_1	4.39	4.18	4.61
$S_2(rain(t))$	β_2	0.92	0.85	0.99
Distance	β_3	-0.19	-0.23	-0.15
Measurement error	σ	0.71	0.70	0.71
Time to event model				
Age	θ_1	0.01	-0.03	0.05
Gender	θ_2	-0.05	-0.21	0.13
Association main effect	α_1	0.12	0.04	0.19
Association interaction	α_2	0.26	0.16	0.36
Hyper-parameters				
Penalty	λ	0.004	0.002	0.008
Random effect covariance	$D_{1,1}$	24.25	22.81	25.80
Random effect covariance	$D_{2,1}$	0.95	0.62	1.28
Random effect covariance	$D_{3,1}$	-0.73	-0.99	-0.47
Random effect covariance	$D_{2,2}$	2.16	2.01	2.30
Random effect covariance	$D_{3,2}$	0.81	0.72	0.91
Random effect covariance	$D_{3,3}$	1.41	1.32	1.51
DIC		403463.5		

D_{ij} denote the ij -element of the covariance matrix for the random effects. Here we use a two week window to define the incidence $I_{k(t)}(t)$. Only rain is used as weather related covariate

risk and seasons are significantly associated. Temperature is negatively correlated to mosquito abundance. This can be explained by the increased evaporation of local water basins and it become less favourable for mosquitoes larvae to survive [8]. Humidity had a marginal role in our model, though it was positively associated to mosquito abundance.

Our results allow quantifying and comparing the contributions to malaria risk of various factors. We found that one additional captured mosquito in the sentinel household increased the relative of risk of malaria by 15%. It is of course not easy to interpret the abundance measure used in our data beyond this study, as it is difficult to say what a unit increase of our abundance measure actually means generally. To try to get an idea of this, we observe that in each village, the captured mosquito numbered from zero to 150 per week per village, and were on average 3.2. This gives an indication of how to understand the mentioned 15% increase.

In this study we used as a measure of the village-wise active malaria incidence the number of new children cases in the previous three weeks, as three weeks is the typical length of a malaria episode, during which the parasite is transmittable [20]. We estimate an increase of the relative

risk of malaria to be 36% for a unit increase of the product of the abundance times the incidence. This can happen for example if both the mosquito count and the incidence increase by one, but there are of course many others combination (like 0.5 and 2, respectively). Interestingly, the interaction between mosquito abundance and incidence has a larger effect, twice as large, on the hazard as compared to abundance alone. It is expected that this interaction has a larger effect than abundance, as both the parasite and the mosquito need to be present for malaria to spread; however the quantification of the relative risks is interesting.

We tested if other reduced models could perform better than the model with temperature, humidity and rain as weather related covariates and distance to the dam as covariate. We computed the DIC measure to compare models, and found that the full models should be preferred. In particular, the distance appears to be an important covariate in the longitudinal model component, supporting the correct way of using such a distance in modelling malaria risk. Models using the reduced definition of measure of the village-wise active malaria incidence to just the last two weeks (instead than three), were not better. We prefer a three weeks window, due to the three week infectiousness of malaria patients. But if we use the results of that analysis, we found that one additional captured mosquito in the household increased the relative risk of malaria by 15%. The increase of the relative risk of malaria is now 36% for a unit increase of the product of the abundance times the incidence.

The importance of the interaction between abundance and incidence leads us to the hypothesis that preventive intervention could advantageously target the infective population, for example by isolation from the rest of the household during disease activity, in addition to mosquito control, which is the main mean of control today.

The construction of dams in Ethiopia has been documented to result in increases in vector populations by creating mosquito habitats [11, 21, 22]. Several studies have investigated the effect of dams on malaria prevalence and incidence [23–25]. In general, dams can have profound effects on the survival, density, and distribution of disease vectors and parasites by altering the local ecology and habitats, and the altered vector/parasite ecology modifies the transmission of vector borne diseases such as malaria and its incidence [26]. Kibret et al. (2009) investigated the association of Koka reservoir and *Anopheles* mosquito density and malaria risk. Accordingly, higher *Anopheles arabiensis* density and malaria incidence was reported from villages in the vicinity of the reservoir. The construction and operation of dams in Senegal River increased anopheline densities, malaria transmission intensity and prevalence was higher in villages closer to dams than in those farther away [27]. A study in Ethiopia also showed

that children living in close proximity to the reservoir created by the newly constructed dam were at a greater risk of *Plasmodium* infection than children living further away [28]. For the Gilgel Gibe cohort, the distance to the dam has not been previously found significant. In our joint model, the distance is a significant factor in the longitudinal mosquito abundance model, and through the coupling of this model to the time to malaria model, it affects the risk of malaria.

Indeed, distance to the dam has a significant effect on the longitudinal mosquito abundance component of the joint model. We can compare the effect of being closer to water and of other risk factors. For example, we showed that for a household to move 1km closer to the water, is (on average) equivalent to capturing 0.13 additional mosquito, which in turn increases the relative risk of malaria by 5.7%. Significant association of mosquito abundance with distance to the dam shore improves the results in [10]: we can now show that the distance has an indirect effect on the malaria infection through the mosquito abundance since the breeding takes place in a water body.

We found that the parameter related to the construction of the houses was not estimated significantly off from zero, despite the 95% credible interval was covering a mostly positive range. This tendency confirms what found in [29], namely that people living in Mozambique in houses with grass or thatch roofs had a greater risk of malaria than those living in houses with corrugated iron roofs.

We did not have access to local precipitation data, and instead used a common measure of rain for all villages. We believe that the effect of precipitation on village-wise mosquito abundance would be stronger if we would have precipitation at a finer spatial scale. Future studies should include local measurements of rain, at least weekly. The dam is likely to be important for mosquito reproduction when there are no alternative water basins in and around the villages, as is very often the case in the rainy season in the study area. Therefore we could expect to find distance to the dam to interact with local measures of precipitations, in affecting mosquito abundance. These more precise factors could not be determined with the current data.

The joint model can be used also to predict time to malaria, given forecasts of mosquito abundance and local malaria incidence. While predictions of malaria free survival curves for individual children appeared to have too large credibility bands, we obtained a good accumulated forecasts across all children. However, this experiment was based on actually incurred abundance and incidence and as such it is of limited interest and therefore not reported further in this paper. More work is needed to make the joint model useful for prediction, especially because uncertainties in covariates need to be

incorporated. The simulation of each individual child ahead in time, using the joint model in a generative way, is a possibility for further work.

Conclusions

In this paper we analyzed for the first time the two data components that play a crucial role in the spread of malaria jointly, namely the abundance of the vector and the appearance in time of new malaria cases. We used regional precipitations, temperature and relative humidity as weather covariates, the presence of a corrugated iron roof and the shortest distance to a dam as covariates. The Bayesian joint model appears to be an appropriate method to perform the combined or joint analysis. In this way we can attribute the effect of distance to water and of meteorological covariates to the mosquito vector, while the time to malaria model depends on these covariates only through the abundance of the vector. We found that the interaction between mosquito abundance and incidence plays the key role. Because of the joint analysis, we can compare the contribution to malaria risk of distance to water and mosquito abundance. Further work will explore the possibility to use the joint model to perform prediction of the risk of future infection in the transmission of malaria in the study area.

Abbreviation

MCMC: Markov Chain Monte Carlo

Acknowledgments

Denekew Bitew Belay was supported by a NORAD/NORHED project. The authors thank referees Nicole White and Sama Low-Choy for their many very helpful and precise comments and suggestions.

Funding

Not applicable.

Availability of data and materials

The data sets of this article are published in [10] and can be obtained from the authors of the paper.

Authors' contributions

YGK, LD and DY designed, collected and prepared the data. DBB, YGK, LD and AF designed and scoped the project. DBB, JMG, ATG and AF designed the statistical approach. DBB performed the statistical analysis and produced all results. All authors discussed the results and contributed to the writing of the paper. All authors read and approved the final manuscript.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

The study was reviewed by the World Health Organization/TDR research ethics committee and ethical approval was obtained from the research and ethics committee of Jimma University. Verbal and written signed informed consent was obtained from the mother or caregiver of each child before enrollment of the child in the study.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹School of Mathematical and Statistical Sciences, College of Natural and Computational Science, Hawassa University, Hawassa, Ethiopia. ²Department of Statistics and Operations Research, University of Limpopo, Limpopo, South Africa. ³Oslo Center for Biostatistics and Epidemiology, University of Oslo and Oslo University Hospital, Oslo, Norway. ⁴Department of Medical Laboratory Sciences and Pathology, College of Health Sciences, Jimma University, Jimma, Ethiopia. ⁵Department of Comparative Physiology and Biometrics, Ghent University, Ghent, Belgium.

Received: 16 November 2016 Accepted: 28 May 2017

Published online: 12 June 2017

References

1. Organization WH, et al. World Malaria Report 2015. Geneva: WHO; 2015.
2. Beier JC. Malaria parasite development in mosquitoes. *Ann Rev Entomol*. 1998;43(1):519–43.
3. Keiser J, Singer BH, Utzinger J. Reducing the burden of malaria in different eco-epidemiological settings with environmental management: a systematic review. *Lancet Infect Dis*. 2005;5(11):695–708.
4. Yewhalaw D, Kassahun W, Woldemichael K, Tushune K, Sudaker M, Kaba D, Duchateau L, Van Bortel W, Speybroeck N. The influence of the gilgel-gibe hydroelectric dam in ethiopia on caregivers' knowledge, perceptions and health-seeking behaviour towards childhood malaria. *Malaria J*. 2010;9(1):1.
5. Himeidan Y, Malik EM, Adam I. Epidemiology and seasonal pattern of malaria in an irrigated area of eastern sudan. *Am J Inf Dis*. 2005;1:75–8.
6. Ferede G, Worku A, Getaneh A, Ahmed A, Haile T, Abdu Y, Tessema B, Wondimeneh Y, Alemu A. Prevalence of malaria from blood smears examination: a seven-year retrospective study from metema hospital, northwest ethiopia. *Malaria Res Treat*. 2013;2013:1–5.
7. Adugna A. Malaria in ethiopia. 2011. Ethiopian Demography And Health Organisation, Addis Ababa.
8. Loha E, Lindtjorn B. Model variations in predicting incidence of plasmodium falciparum malaria using 1998–2007 morbidity and meteorological data from south ethiopia. *Malaria J*. 2010;9(1):1.
9. Lunde TM, Balkew M, Korecha D, Gebre-Michael T, Massebo F, Sorteberg A, Lindtjorn B. A dynamic model of some malaria-transmitting anopheline mosquitoes of the afrotropical region. ii. validation of species distribution and seasonal variations. *Malaria J*. 2013;12(1):1.
10. Getachew Y, Janssen P, Yewhalaw D, Speybroeck N, Duchateau L. Coping with time and space in modelling malaria incidence: a comparison of survival and count regression models. *Stat Med*. 2013;32(18):3224–33.
11. Yewhalaw D, Getachew Y, Tushune K, Kassahun W, Duchateau L, Speybroeck N, et al. The effect of dams and seasons on malaria incidence and anopheles abundance in ethiopia. *BMC Infect Dis*. 2013;13(1):1.
12. Rizopoulos D. The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. 2014. arXiv preprint arXiv:1404.7625.
13. Rizopoulos D. Joint Models for Longitudinal and Time-to-event Data: With Applications in R. Boca Raton: CRC Press; 2012, pp. 97–170.
14. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J Royal Stat Soc Ser B (Stat Method)*. 2002;64(4):583–639.
15. Eilers PH, Marx BD. Flexible smoothing with b-splines and penalties. *Stat Sci*. 1996;11(2):89–102.
16. Cox DR. Regression models and life-tables. In: Breakthroughs in Statistics. New York: Springer; 1992. p. 527–41.
17. Lang S, Brezger A. Bayesian p-splines. *J Comput Graph Stat*. 2004;13(1):183–212.
18. Barnard R, McCulloch R, Meng XL. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Stat Sinic*. 2000;1281–311.
19. Gelman A, et al. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Anal*. 2006;1(3):515–34.

20. Centers for disease control and prevention. <https://www.cdc.gov/malaria/about/disease.html>. Accessed 05 Jan 2016.
21. Kibret S, McCartney M, Lautze J, Jayasinghe G. Malaria Transmission in the Vicinity of Impounded Water: Evidence from the Koka Reservoir, Ethiopia. Vol. 132. Colombo: IWMI; 2009, pp. 1–33.
22. Lautze J, McCartney M, Kirshen P, Olana D, Jayasinghe G, Spielman A. Effect of a large dam on malaria risk: the koka reservoir in ethiopia. *Trop Med Int Health*. 2007;12(8):982–9.
23. Keiser J, de Castro MC, Maltese MF, Bos R, Tanner M, Singer BH, Utzinger J. Effect of irrigation and large dams on the burden of malaria on a global and regional scale. *Am J Trop Med Hyg*. 2005;72(4):392–406.
24. Singh N, Mehra R, Sharma V. Malaria and the narmada-river development in india: a case study of the bargi dam. *Ann Trop Med Parasitol*. 1999;93(5):477–88.
25. Ghebreyesus TA, Haile M, Witten KH, Getachew A, Yohannes AM, Yohannes M, Teklehaimanot HD, Lindsay SW, Byass P. Incidence of malaria among children living near dams in northern ethiopia: community based incidence survey. *Bmj*. 1999;319(7211):663–6.
26. Yewhalaw D, Hamels S, Getachew Y, Torgerson P, Anagnostou M, Legesse W, Kloos H, Duchateau L, Speybroeck N. Water resource developments in ethiopia: potential benefits and negative impacts on the environment, vector-borne diseases, and food security. *Environ Rev*. 2014;22(4):364–71.
27. Dia I, Konate L, Samb B, Sarr JB, Diop A, Rogerie F, Faye M, Riveau G, Remoue F, Diallo M, et al. Bionomics of malaria vectors and relationship with malaria transmission and epidemiology in three physiographic zones in the senegal river basin. *Acta tropica*. 2008;105(2):145–53.
28. Yewhalaw D, Legesse W, Van Bortel W, Gebre-Selassie S, Kloos H, Duchateau L, Speybroeck N. Malaria and water resource development: the case of gilgel-gibe hydroelectric dam in ethiopia. *Malaria J*. 2009;8(1):1.
29. Charlwood JD, Tomás EV, Bragança M, Cuamba N, Alifrangis M, Stanton M. Malaria prevalence and incidence in an isolated, meso-endemic area of mozambique. *PeerJ*. 2015;3:1370.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

